Biol Philos (2014) 29:181–195
DOI 10.1007/s10539-013-9415-x

Fitness, inclusive fitness, and optimization

Laurent Lehmann · François Rousset

Accepted: 23 December 2013 / Published online: 25 January 2014

Abstract Individual-as-maximizing agent analogies result in a simple understanding of the functioning of the biological world. Identifying the conditions under which individuals can be regarded as fitness maximizing agents is thus of considerable interest to biologists. Here, we compare different concepts of fitness maximization, and discuss within a single framework the relationship between Hamilton's (J Theor Biol 7:1–16, 1964) model of social interactions, Grafen's (J Evol Biol 20:1243–1254, 2007a) formal Darwinism project, and the idea of evolutionary stable strategies. We distinguish cases where phenotypic effects are additive separable or not, the latter not being covered by Grafen's analysis. In both cases it is possible to define a maximand, in the form of an objective function $\phi(z)$, whose argument is the phenotype of an individual and whose derivative is proportional to Hamilton's inclusive fitness effect. However, this maximand can be identified with the expression for fecundity or fitness only in the case of additive separable phenotypic effects, making individual-as-maximizing agent analogies unattractive (although formally correct) under general situations of social interactions. We also feel that there is an inconsistency in Grafen's characterization of the solution of his maximization program by use of inclusive fitness arguments. His results are in conflict with those on evolutionary stable strategies obtained by applying inclusive fitness theory, and can be repaired only by changing the definition of the problem.

Keywords Fitness · Inclusive fitness · Maximization · Optimization program · Game theory · Dynamic sufficiency

L. Lehmann and F. Rousset contributed equally to this work.

L. Lehmann (✉)
Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland
e-mail: Laurent.Lehmann@unil.ch

F. Rousset
CNRS, Institut des Sciences de l'évolution, Université Montpellier II, Paris, France

Introduction

It is plausible that ant colonies adjust their collective behaviour to maximize food intake. Plants may regulate biochemical cycles to maximize photosynthesis under different constraints of pH or water and carbon dioxide availability. A bird wing shape may be built to maximize aerodynamic efficiency for different kinds of flight. The application of such optimality considerations to understand the form, physiology, and behavior of organisms has often enhanced the understanding of biological processes (Maynard Smith 1982; Dawkins 1982; Parker and Maynard Smith 1990).

The theory of natural selection itself was originally developed as a way of explaining the perceived optimal fit of organisms to their physical environment. However, much of the attraction of natural selection also stems from its ability to understand processes involving social interactions, from iconic examples of sex ratio evolution to the analysis of conflicts at all levels of biological organization (e.g., Maynard Smith and Szathmary 1995; Bourke 2011). A long-standing question in evolutionary biology is the extent to which natural selection leads individuals to behave *as if* they maximize a common measure, "fitness", of fit to the environment in all these different cases, and then what this fitness means.

In a series of papers, Grafen (2002, 2006, 2008) appears to have constructed general results in the form of individual-as-maximizing agent analogies and describes these results as a general formal statement of Darwin's theory of natural selection. What he appears to be after is the formal maximization of a function $\phi(z)$ with respect to an individual phenotype z. The problem then is to find the appropriate function ϕ.

In order to identify this individual maximand, Grafen describes, in particular with his concept of "no potential for selection", the mathematical characterization of a concept of evolutionary stability ("no possible mutant would spread", Grafen 2008, p. 425). This is supposed to go beyond more traditional concepts from evolutionary game theory (Maynard Smith 1982; Eshel 1983) in two directions. First, it is based on explicit population genetic considerations and, second, the characterization must apply to an arbitrary genetic makeup of a given parental population.

Many steps of Grafen's argument are sound. For instance, his stressing of the importance of having a consistent usage of the word "fitness", and whose arithmetic mean can be applied to understand the effect of natural selection even in cases where biologists feel a need to consider geometric means (Grafen 1999, 2000). But we are skeptical about the biological importance of the results reached on maximization, for two reasons. First, the exact significance of Grafen's (2006) results on the maximization of inclusive fitness may easily be missed. They require the assumption of additive separable phenotypic effects on fitness (ruling out phenotypic interactions), which is stronger than additive separable genetic effects. Hamilton's (1964) model still works with the latter assumption and thus applies to phenotypic interactions (the rule in the presence of social interactions). Second, Grafen's mathematical characterization of fitness maximization does not always appear consistent with well-established population genetic and game theoretical

considerations. In particular, it appears to us as inconsistent with inclusive fitness theory.

In this commentary, we develop the above points. We discuss the validity of different concepts of fitness maximization in Hamilton's (1964) model of social interactions, in the formal Darwinism project of Grafen (2014), and how these relate to each other and to the idea of evolutionary stability (Maynard Smith 1982; Eshel 1983). This paper is organized as follows. (1) We start by discussing fitness maximization in population genetics. (2) We analyze fitness maximization in Hamilton's (1964) model of social interactions, where candidate maximands depend on gene frequency. With the possible exception of our comparison of partial and total changes in fecundity under this model, our analyses are not new, but are profitably set in a common framework. (3) We relate Hamilton's model to the concept of evolutionary stability, where candidate maximands now depend on phenotypes. We then compare maximands under two different altruism models, one involving additive separable phenotypic effects and the other not. While we show that individual-as-maximizing agent analogies still appear formally correct in the latter case, they generally do not provide new biological insights. (4) In light of these results, we call for several clarifications in Grafen's arguments in the case of social interactions.

Fitness and optimization

Fitness

For simplicity, we assume throughout that evolution occurs in a haploid population of constant size N without any class structure. We denote by w_i the fitness of individual i in this population, and follow Hamilton's (1964) words in defining this as the number of offspring in a daughter generation that descend from individual i reproducing in a given parental generation. Because population size is constant, mean fitness is equal to one ($\bar{w} = \sum_i w_i/N = 1$). If a trait or characteristic with value y_i in individual i is transmitted identically from parent to offspring, the change $\Delta\bar{y}$ in the average $\bar{y} = \sum_i y_i/N$ of that trait over one generation can then be written as

$$\Delta\bar{y} = \sum_i y_i w_i/N - \sum_i y_i/N \sum_i w_i/N = \mathrm{Cov}(y_i, w_i). \quad (1)$$

This is a particularly simple formulation of the classic result of Price (1970), which is in agreement with Grafen's (2008) "simplest model", and where the covariance is taken over all population members.

It is tempting to set $y_i = w_i$ in Eq. (1), which gives the change in mean fitness as the covariance in fitness: $\Delta\bar{w} = \mathrm{Cov}(w_i, w_i) = \mathrm{Var}(w_i)$, but mean fitness should not change over generations in a population of constant size ($\Delta\bar{w} = 0$). What goes wrong here? The answer is that the w_i's are not identically transmitted from parent to offspring. This fact may be framed in terms of Fisher's (1930) so-called Fundamental Theorem of Natural Selection, as explained by Price (1972).

According to this interpretation, $\mathrm{Var}(w_i)$ only represents a partial change in the average fitness $\bar{w}$, attributed by Fisher to "natural selection". But the total change in $\bar{w}$ is also affected by changes in the "environment", which, in a model of social interactions, encapsulates the genetic effects (and thus behavior) of other individuals in the population. The change in the environment thus includes changes in the genetic composition of the population as the result of natural selection. The partial change $\mathrm{Var}(w_i)$ isolates that part of the mean change about which something can be said independently of what is know about the parental generation (Ewens 2011, p. 169), but this is exactly counter-balanced by the change in the genetic environment from the parental to offspring generation.

The idea that natural selection always results in such a simple concept of adaptation as an increase in mean fitness (the "mean fitness program" in the words of Grafen 2008, p. 424) has been criticized and assessed in population genetics (Moran 1964; Ewens 2004, 2011) and evolutionary game theory (Mylius and Diekmann 1995; Metz et al. 2008). Yet Hamilton (1964) attempted to show that "inclusive fitness" would always increase. Hamilton's result may thus appear as an instantiation of the mean fitness program. However, we now show that Hamilton's 1964 result is an instantiation of the partial change in mean fitness result. In so doing, we will not use Hamilton's notations, but follow his line of arguments applied to a simple example. Hence, all results presented in the next section can be seen as special case of Hamilton's (1964) model.

Social interactions

Partial change in fitness

Hamilton (1964) assumed a population without spatial structure, with discrete and non-overlapping generations, and where the fitness $w_i = f_i/\bar{f}$ of individual i depends on the average fecundity $\bar{f}$ in the population. In a model with only two alleles, the fecundity f_i of individual i may depend not only on the frequency p_i by which it carries the mutant allele, but also on the fraction $p_{\mathrm{n},i}$ of neighbours it interacts with that carry the mutant. The fecundity of individual i can then be written as $f_i = f_{\mathrm{b}}(1 - Cp_i + Bp_{\mathrm{n},i})$ for some baseline fecundity f_{b}, fecundity cost C of expressing the mutant allele, and fecundity benefit B received from neighbors that express the mutant.

Fecundity f_i is not identically transmitted across generations, because in general $p_{\mathrm{n},i}$ is not identically inherited from parent to offspring ($p_{\mathrm{n},i}$ is part of the "environment"). However, there are two further steps to Hamilton's reasoning. First, the population is very large, each allele is in many copies, and random fluctuations in $p_{\mathrm{n},i}$ average out over all gene copies i of a given allelic type. Second, the expected $p_{\mathrm{n},i}$ for individual i takes the form $Rp_i + (1 - R)\bar{p}$ in terms of the mutant allele frequency $\bar{p}$ in the total population (Hamilton 1964, p. 35), which can be interpreted as saying that a fraction R of gene copies in neighbors are identical to those of individual i, while a complementary fraction are mutant gene copies according to its frequency in the population. For a mutant, this entails expected

frequency $R + (1 - R)\bar{p}$ of interactions with other mutants and expected fecundity $f_b(1 - C + [R + (1 - R)\bar{p}]B)$, while for a resident (or wild-type) it entails frequency $(1 - R)\bar{p}$ of interactions with mutants and expected fecundity $f_b(1 + (1 - R)\bar{p}B)$.

The difference between the two expected fecundities is $f_b(- C + RB)$, and, the average mutant frequency change in the population can be written as

$$\Delta\bar{p} = \bar{p}(1 - \bar{p})(-C + RB)f_b/\bar{f}. \tag{2}$$

Because selection acts on fecundity differences in this model, Hamilton showed that the change in allele frequency in the population is *as if* the fecundity of individual i is

$$f_{a,i} = f_b[1 + p_i(-C + RB)], \tag{3}$$

which is a value that can be associated to each gene copy (equal to f_b for a wild-type and $f_b(1 - C + RB)$ for a mutant). Hamilton (1964, p. 6) called this value "inclusive fitness", a semantic choice consistent with the usage in the population genetic literature that inspired him, but is inconsistent with his own verbal definition of "fitness" as a number of adult offspring (Hamilton 1964, p. 1), which matches w_i defined above. In order to avoid such semantic inconsistencies, and further semantic difficulties that arise in models of spatially structured population (where regulation is local), we prefer to call this value "fecundity asif" to emphasize the precise interpretation of Eq. (3).

With the definition of fecundity asif, the expected fecundity of individual i can be written as

$$\mathrm{E}[f_i] = f_{a,i} + f_b B(1 - R)\bar{p}, \tag{4}$$

which is the sum of fecundity asif and a remainder term depending on population allele frequency. The total change in fecundity asif is then given by $\Delta\bar{f}_a = \mathrm{Cov}(f_{a,i}, w_i) = \mathrm{Cov}\left(f_{a,i}, [f_{a,i} + f_b(1 - R)\bar{p}B]/\bar{f}\right) = \mathrm{Var}(f_{a,i})/\bar{f}$. Using the explicit expression for $f_{a,i}$ and the identity $\mathrm{Cov}(p_i, p_i) = \mathrm{Var}(p_i) = \bar{p}(1 - \bar{p})$ produces

$$\Delta\bar{f}_a = \bar{p}(1 - \bar{p})(-C + RB)^2 f_b^2/\bar{f}. \tag{5}$$

Therefore, the fecundity asif always increases in the population as long as allele frequency change occurs, and the same argument can be made for the fitness asif $\bar{w}_a$, defined in terms of fitness w_i by dividing all expressions involving fecundity by $\bar{f}$.

In Hamilton's construction, the mean fecundity asif, $\bar{f}_a = f_b(1 + \bar{p}[-C + RB])$, acts as a so-called potential function, which is increased by evolutionary change. That is, the gradient $\mathrm{d}\bar{f}_a/\mathrm{d}\bar{p} = f_b(-C + RB)$ of the potential points in the direction of the steepest increase in fecundity asif, which is the path taken by allele frequency change:

$$\Delta\bar{p} = \frac{\bar{p}(1 - \bar{p})}{\bar{f}} \frac{\mathrm{df}_a}{\mathrm{d}\bar{p}}. \tag{6}$$

As proven, this result does not imply that fecundity is maximized under biological scenarios involving social interactions. In fact, Eq. (5) only provides the partial change in fecundity due to changes in allele frequencies in the population,

but given the mean fecundity in the parental population. This mean fecundity also changes as the result of allele frequency change. The key relationship is here Eq. (4), where it is seen that differences among alleles in fecundity asif equal differences in expected fecundity f_i, but that the fecundity asif differs from fecundity by a function of allele frequency which is Hamilton's (1964, p. 6) diluting effect.

Total change in fitness

In order to obtain the total change, it suffices to note that each mutant allele imparts a total cost $-C$ and a total benefit B on the relative fecundity of the population. The overall effect of each mutant on relative population fecundity is $B - C$. This is an exactly transmitted property of each mutant allele. Hence, the total change in average population fecundity depends on the extent to which allele frequency change alters this value: $\Delta\bar{f} = f_b(B - C)\Delta\bar{p}$, and using Eq. (2) produces

$$\Delta\bar{f} = \bar{p}(1 - \bar{p})(B - C)(-C + RB)f_b^2/\bar{f}. \quad (7)$$

This shows that average fecundity will decrease in the population as genes with higher relative fecundity increase in frequency: $-C + RB > 0$, but absolute fecundity decreases ($B - C < 0$); namely, when $B < C < RB < 0$. A "selfish" mutant with positive direct effects but larger negative indirect effects on weakly related neighbours is selected for. The case where a selfish mutant invades despite imparting a negative effect on the whole population ($R \sim 0$) is indeed an intuitive case of this more general result, which also underlies selection-driven population extinction (Matsuda and Abrams 1994).

To sum up, and as claimed by Hamilton, allele frequency changes proceed as if fitness was proportional to $f_b[1 + p_i(- C + RB)]$. The average fecundity asif $\bar{f}_a$ therefore increases in the population as the mutant invades, and the same argument holds for fitness asif. Indeed, Hamilton's argument was that the change of allele frequency due to selection proceeds *as if* individuals were changing behaviour to increase their fitness asif. Hamilton's (1964, 1970) model thus appears analogous to previous works by Wright (1942) and Kingman (1961) to which he refers, and which are instantiations of the "mean fitness program". In effect, Eq. (6) takes the same form as the influential "adaptive topography" equation of Wright (1942). But the analogy holds only as long as one deals with allele frequency changes, but not with changes in fecundity or fitness, as there is nothing in Hamilton's result that prevents these quantities from going down.

Optimization

Continuum of phenotypes

So far, the fecundity asif, $f_{a,i}$, was not considered a function of all the alternative phenotypes that can be expressed by an individual, and therefore not considered as an objective function that can be maximized by varying its behavior over an

arbitrary phenotypic range. But what Grafen is seeking in is his "optimization" papers (Grafen 2002, 2006, 2008) is such an objective function $\phi(z)$, whose argument z is the phenotype of an individual (more precisely that part of the phenotype that vary with genotype holding everything else constant), and that can represent its state, from physiological to informational.

In order to capture the (competitive) fit of an organism to its environment, this maximand must allow one to characterize evolutionary stable strategies in the sense that if all individuals in a population express the phenotypic value $z^\star$ that maximizes the objective function (such that $\phi(z^\star) = \max_{z\in\Phi} \phi(z)$, where Φ is the set of phenotypes), no mutant with a deviant phenotype can invade the population. The non-invadability condition of mutants is captured by the concept of "no potential for selection in relation to the set Φ" in Grafen's work (e.g., Grafen 2008, p. 425).

Can one find such a maximand in the framework based on Hamilton's model described above? In order to answer this question, we write the fecundity cost C and B explicitly in terms of an evolving phenotype, whose range Φ is assumed to be real valued (continuously distributed phenotype). For instance, this phenotype could be the probability of committing self-sacrifice ($\Phi = [0, 1]$). The fecundity of individual i in an altruism model could then be written as

$$f_i = f(z_i, z_n) = f_b(1 - z_i)(1 + \alpha z_n), \tag{8}$$

where α is the increase in fecundity when a focal individual that has not committed self-sacrifice interacts with an altruistic neighbor. This is a standard formulation for an altruism model in the literature (Charlesworth 1978; Frank 1998).

The change in frequency of a mutant allele with phenotype $z + \delta$ in a wild-type population with phenotype z is then given for a mutant with small phenotypic deviation δ by

$$\Delta\bar{p} = \bar{p}(1 - \bar{p})\delta[-C(z) + RB(z)]/f(z, z), \tag{9}$$

where

$$\begin{aligned} -C(z) &= \partial f(z_i, z_n)/\partial z_i|_{z_i=z_n=z} \\ B(z) &= \partial f(z_i, z_n)/\partial z_n|_{z_i=z_n=z}. \end{aligned} \tag{10}$$

For the altruism model, these marginal cost and benefit are $-C(z) = -f_b(1 + \alpha z)$ and $B(z) = f_b(1 - z)\alpha$, respectively. It is in terms of such marginal costs and benefits that Hamilton's (1964) model should be thought of, otherwise the relatedness coefficients would not behave as claimed and would depend on frequency $\bar{p}$ of the mutant. In terms of the marginal cost and benefit, the fecundity asif of individual i is $f_b + p_i[-C(z) + RB(z)]$, its mean is $\bar{f}_a(z, \bar{p}) = f_b + \bar{p}[-C(z) + RB(z)]$ and all the results obtained in the previous section apply *mutatis mutandis*. In particular, the mutant invades a population of wild-types when the selection gradient $S(z) = [\partial\bar{f}_a(z, \bar{p})/\partial\bar{p}]/f(z, z)$ is positive, *as if* individuals were changing their behaviour to maximize their fitness asif. But what about maximization of the fitness asif with respect to z?

Evolutionary potential function

Because the gradient $S(z)$ is of constant sign, invasion of a mutant allele implies its fixation in the population. By successive allelic replacement, the level of altruism in the population will gradually change. For a constant mutation rate and phenotypic variance, the change in phenotype under a trait substitution sequence assumption (e.g., Metz et al. 1996) is proportional to the gradient: $\mathrm{d}z/\mathrm{d}t = kS(z)$, where the constant k of proportionality determines the rate of evolution. This equation for the change in phenotype (which neglects the possibility of stable polymorphism in the population) is the so-called canonical equation of adaptive dynamics (Dieckmann and Law 1996; Champagnat et al. 2006), which can be derived by using the population genetic assumptions behind Hamilton's model (Lehmann 2012). It thus applies to interaction between relatives and provides the direct long-term phenotypic evolution counterpart to the short-term evolutionary model discussed in the last section (Eq. 2).

It is useful to note that the selection gradient on the level of altruism can be interpreted as the gradient of the potential function

$$\phi(z) = \int S(z)\mathrm{d}z, \tag{11}$$

whereby the change of phenotype in the population is

$$\frac{\mathrm{d}z}{\mathrm{d}t} = k\frac{\mathrm{d}\phi(z)}{\mathrm{d}z}. \tag{12}$$

Evolution stops when $\mathrm{d}z/\mathrm{d}t = 0$. This occurs in point z where the selection gradient vanishes: $\mathrm{d}\phi(z)/\mathrm{d}z = S(z) = 0$. It entails no change of allele frequency and thus characterizes a candidate evolutionary stable strategy if $\phi(z)$ is a local maximum, so that no nearby deviant mutant can invade. Thus, if the individuals in the population behave *as if* they were maximizing $\phi(z)$, no nearby deviant mutant can invade. In the altruism model, this entails maximizing

$$\phi(z) = R\log(1+\alpha z) - \log(1-z) \tag{13}$$

and expressing level of altruism $z^\star = (\alpha R - 1)/[\alpha(1+R)]$.

In the absence of social interactions, the fecundity of an individual depends only on its own phenotype ($f_i = f(z_i)$ does not depend on z_n, nor on z). Then, the inclusive fitness effect can be written as $S(z) = \partial\bar{f}_a(z,\bar{p})/\partial\bar{p} = [\mathrm{d}f(z)/\mathrm{d}z]/f(z)$ and we can take $\phi(z) = \log f(z)$, which is the logarithm of fecundity. If the individuals in a population then behave *as if* they were maximizing $\log f(z)$ or simply $f(z)$, no mutant in relation to the whole set Φ can invade; and we can even remove the term "nearby mutant" in this case. The maximand thus allows one to characterize evolutionary stable phenotypes and provides an intuitive individual-as-maximizing-agent analogy.

In the presence of social interactions, the fecundity of an individual no longer depends only on its own phenotype and it is no longer clear what the maximand $\phi(z)$ really represent biologically. Indeed, in the altruism model given above (Eq. 13), neither $\log(1-z)$ nor $\log(1+\alpha z)$ have a clear biological interpretation in terms

of vital rates of actors and/or recipients. Further, relatedness itself may depend on the evolving phenotype, as occurs when dispersal is the evolving phenotype (e.g., Frank 1998; Rousset 2004). In these cases, the individual-as-maximizing-agent analogy of $\phi(z)$ becomes less seductive, even if formally correct. But $\phi(z)$ still retains a biological effect, as it determines the distribution of phenotypes at a mutation-selection-drift equilibrium, and thus arises naturally in a model where the continuum of possible phenotypes in Φ are explicitly taken into account under arbitrary kinds of asymmetric interactions and environmental or demographic stochasticity (Lehmann 2012).

Additive separable phenotypic effects

There is, nevertheless, a case where the evolutionary potential function takes a clear biological interpretation in the presence of social interactions. Consider the altruism model where the fecundity of individual i is written as

$$f(z_i, z_n) = f_b - \gamma(z_i) + \beta(z_n) \tag{14}$$

for some cost function γ and benefit function β entailing additive separable phenotypic effects on fecundity. This is an alternative formulation to Eq. (8) of an altruism model, and also appears in the literature (Frank 1998; Lion and Gandon 2009).

For this model the selection gradient is $S(z) = [-\mathrm{d}\gamma(z)/\mathrm{d}z + R\mathrm{d}\beta(z)/\mathrm{d}z]/f(z,z)$ and one can define the evolutionary potential

$$\phi_s(z) = -\gamma(z) + R\beta(z), \tag{15}$$

which allows us to write the selection gradient as

$$S(z) = \frac{1}{f(z,z)} \frac{\mathrm{d}\phi_s(z)}{\mathrm{d}z}. \tag{16}$$

If the individuals in the population behave *as if* they were maximizing ϕ_s, no nearby mutant can invade, and this result applies more generally; namely, to all cases where $\phi(z)$ applies and where social interactions result in a fecundity function (or fitness function) that is additive separable.

Equation (15) sums up the relatedness weighted cost and benefit of social interactions, and is sometimes used as "inclusive fitness" in the literature, in particular in reproductive skew or tug-of-war models (e.g., Johnstone et al. 1999). This definition departs from the initial conception of Hamilton (Eq. 3) in a crucial way. In effect, his equations also apply to all cases where gene action is additive under weak selection (so that there are additive effects on fitness stemming from differences in behaviour between competing alleles, e.g., Taylor 1989; Rousset 2004). For instance, they apply in the altruism model (Eq. 8), where phenotypic interactions are not additive separable, but weak selection entails additive gene action. Many other applications of Hamilton's rule involve such phenotypic interactions, e.g., phenotype-matching kin recognition (Reeve 1989) or the evolution of sex-ratio, over-exploitation of resources, or policing (Frank 1998; Wenseleers et al. 2010).

Grafen's program

We now discuss the results of Grafen's "optimization" papers (Grafen 2002, 2006, 2008) in the light of the inclusive fitness and game theoretic results introduced above. One of the main reason that we presented these results is that we failed to find an unambiguous relationship between Grafen's concepts of fitness and maximization, and those used by Hamilton (1964) and in classical ESS calculations with and without relatives (e.g., Parker and Maynard Smith 1990; Frank 1998; McNamara et al. 2001). Therefore, our aim in the forthcoming section is not so much to discuss all of Grafen's claims about maximization (which we may not fully understand), but rather to give elements that should help readers to evaluate future clarifications of these claims. A major issue is to find the function to which Grafen's program applies. In our understanding, a function of at least two variables, such as the fitness function, does not fit with Grafen characterization of the maximand. As emphasized by Grafen (2008, p. 423), there is not even a concept of population involved in the definition of the maximand, so that the phenotypes of different individuals cannot be considered in this definition. This strict concept of maximization thus excludes the concept of "best response" (Mas-Colell et al. 1995, p. 242), that is, maximization with respect to one argument by holding the others constant, which actually often underlies ESS calculations (e.g., Parker and Maynard Smith 1990; Mylius and Diekmann 1995). While the selection gradient is a function of a single variable, it takes a value of zero at an ESS point and is thus not the required maximand. The remaining candidate encountered above for a maximand is the evolutionary potential function, from which we suggest that the maximands proposed by Grafen (2002, 2006, 2008) can be retrieved.

Asocial worlds

In his first paper demonstrating the existence of a maximand, Grafen (2002) assumes no social interactions. The fecundity (or survival) of an individual thus depends only on its own phenotype ($f_i = f(z_i)$). In this case, we saw that if individuals behave *as if* they maximize $f(z)$, one obtains a characterization of evolutionary stability so that no mutant can invade ("no potential for selection in relation to the set Φ"). This has been noticed before and the maximands $\phi(z) = f(z)$ (or $\phi(z) = \log f(z)$) form the basis of much of behavioral ecology in the absence of social interactions (e.g., marginal value theorem, Charnov 1976) and life-history evolution (e.g., semelparity vs. iteroparity, Stearns 1992).

Grafen (2002, 2008) proves the result that organisms may be regarded as fecundity/survival maximizers under conditions more general than assumed above, and extends it to an arbitrary ploidy, number of loci, and uncertainty. As emphasized by Grafen himself (e.g., Grafen 2007a, p. 1248), it is not a conclusion that natural selection will necessarily lead to optimization under such conditions, and, importantly, the classical population genetic restriction to optimization noted in the section "Fitness" (e.g., Moran 1964) still apply to his model. But by emphasizing the phenotypic optimization and population genetic parts of the same model, Grafen provides a more detailed justification of previous models assuming

optimality, in particular in behavioral ecology (e.g., Charnov 1976), and his characterization is sufficient to determine the candidate endpoints of the evolutionary process when genetic constraints are ignored.

In effect, many parts of an organism appear *as if* they have been optimally engineered. From molecular motors and pumps to swim bladders and the eye, there are many morphological and behavioral traits that seem to ideally fit the prevailing environmental conditions. The behavior of individuals from other species is taken as constant in the maximand, so that the immune system, spider webs, or the ability to evade predators, all fall to some extent into the ambit of Grafen's model.

Social worlds

Additive separable phenotypic effects

In the presence of social interactions, however, the behaviors of individuals often do no look as if there were at their best. Overexploitation of resources, nepotism, and conflicts at various levels of social organization do not carry the hallmarks of optimal design. Indeed, when the fecundity (or survival) of an individual depends on the behavior of conspecifics, the Pareto optima of a game are often not Nash equilibria.

In his (2006) paper, Grafen extends his (2002) results and claims to show that in the presence of social interactions individuals behave as if they maximize their inclusive fitness effect (the maximand is verbally defined as such by Grafen 2006, p. 552 and given explicitly by Eqs. 8–9). This is inconsistent with the results discussed above, since the inclusive fitness effect is proportional to $S(z) = [-C(z) + R\,B(z)]/f(z, z)$ and takes a value of zero at an interior candidate evolutionary stable state [following the definition of Hamilton (1964, p. 6), the inclusive fitness effect is $\delta S(z)$]. This inconsistency is resolved by noting that Grafen's construction of inclusive fitness is in terms of additive separable phenotypic effects (e.g., Eq. 15, Grafen personal communication). Hence, the maximand should be $\phi_s(z)$; that is, Eq. (15) or its generalization to asymmetric interactions and/or stochastic demographies and environments. As discussed above, this is but a special case of the domain of application of Hamilton's model, and this maximand is not the inclusive fitness effect *per se*, but still bears a simple enough relationship to it for the two to be often confounded.

Dynamic sufficiency

We are actually further puzzled by Grafen's (2006) treatment of the inclusive fitness effect, because it seems that the kind of partial change Grafen is after is so partial that it actually does not contain any inclusive fitness effects in its formulation. If so, his characterization of "no scope for selection" and "no potential for selection" will appear removed from the evolutionary stability considerations such as those outlined above. If not, his results need to be reconciled with the following considerations.

Grafen (2006) compares changes in the transmission of a gene copy when a single individual in a population of wild-types switches to the expression of a

mutant allele. The result (Grafen 2006; p. 553, eq. 10) may then be seen to depend only on the direct effect of the individual on its fitness, not on its effects on related neighbours. Namely, on $-C(z)$ [or $-\gamma(z)$] rather than on $-C(z) + RB(z)$ [or $-\gamma(z) + R\beta(z)$] in our altruism example. This result and its derivation depart from Hamilton's logic in a crucial way. In the latter logic, the role of the inclusive fitness effect in determining allele frequency change is recovered in the following comparison: when a gene copy switches from one behaviour to another, the behaviour expressed by any gene copy related by R to the first one should also be altered with probability R. In other words, considering the fate of a single switch in behaviour over one episode of reproduction is not indicative of the direction of selection on a mutant in the presence of interactions between relatives, and thereby decoupled from any consideration of evolutionary stability.

It is plausible that Grafen's (2006) result for the change in allele frequency (his Eq. 10) also applies to the case where there are several mutants. In this case, however, the inclusive fitness effect in this equation will depend on allele frequency, since relatedness depends on the distribution of allele frequencies in the population (Grafen 2006, Eq. 3). Thus, the maximand will depend on the setups of the two parental populations that are compared. But in Grafen's approach, what determines this setup is not considered. Indeed, since the model is not dynamically sufficient, as emphasized by Grafen (2008, p. 431; 2007a, p. 1247), we are not in a position to say anything about the parental population setup. We do not see, however, how one can provide a formal foundation to phenotypic optimization by letting the relatedness coefficient, and thus the maximand, vary with the parental population setup. This criticism is not inspired by the classical population genetic counterexamples to optimization found when dynamics sufficiency is taken into account. But by the fact that Grafen's characterization of "no potential for selection in relation to the set Φ" is then at variance with the usual notion of non-invadability, which is a procedure that allows one to determine the candidate endpoints of an evolutionary process (e.g., Parker and Maynard Smith 1990; McNamara et al. 2001). This should take into account the likelihoods of various population configurations in order to make predictions about the behaviors that are likely to be observed in a population.

Indeed, Hamilton's argument leading to the expression for change in fecundity (or survival) in terms of relatedness (Eq. 4) shows that the parental setup matters and must be chosen in a biological meaningful way, rather than considered as an arbitrary given. In the simplest population genetic scenario without social interactions, the change in allele frequency is of the form $p(1-p)s$ for some constant selection coefficient s (Crow and Kimura 1970; Gillespie 2004), and thus the direction of selection is given by s irrespective of p: we do not need to care about making dynamically sufficient claims about p (this is one of the reasons why Grafen's characterization of the asocial case is relevant and also pertains to long-term evolution). In Hamilton's scenario, the change in allele frequency is similar in form, $p(1-p)S(z)$, but, importantly, this result rests on the average genetic structure in the parental population given p. In a more formal analysis of Hamilton's model, this genetic structure may be obtained as the result of a dynamically sufficient analysis of probabilities of identity between pairs of genes (Rousset 2004),

so that $S(z)$ (and $\phi(z)$ in Eq. 11) encapsulate the likelihoods of various population configurations. Hence, the maximization of the evolutionary potential $\phi(z)$ makes prediction about the behaviors that are likely to be observed in the population in the long run. By contrast, Grafen's (2006) measure of relatedness (his Eq. 3), and then his maximand, depend on the the realized distribution of allele frequencies in a given parental population, so that it is not clear how its maximization pertains to the behaviors to be observed in the population.

Under more general biological scenarios, the parental genetic structure may be more difficult to characterize. One way of circumventing this problem it to derive results for fixation probabilities (rather than allele frequency change), obtained as integrals of some functions over sample paths of a stochastic process determining likely properties of the parental population (Rousset 2003, 2004; Lehmann 2007; Lessard and Ladret 2007). In such an approach, one still has to say something about the likely state of the parental population in order to obtain a measure of the direction of selection on a mutant and then to determine the evolutionary potential individuals may appear to be maximizing as the result of natural selection.

Metamodel?

By contrast to the previous arguments, Grafen's (2006) computation gives mutant spread only under conditions decoupled from a consideration of evolutionary stability. We thus have difficulties to reconcile this approach with Grafen's claim in his (target review) that his project has the status of a meta model (Grafen 2014), by which the processes going on in other models can be understood. For instance, he claims that the paper of Grafen (2007b) follows the assumptions of Grafen (2006) and allowed to show that the results of Ohtsuki et al. (2006) for games on graphs can be understood in terms of inclusive fitness effects. To do this, however, Grafen (2007b) in his Appendix does not apply his characterization of solutions of the optimization program, but precisely the dynamically sufficient approach on probabilities of identity mentioned above.

Conclusion

Are there general conditions where individuals can be regarded as fitness maximizing agents? Hamilton (1964) showed that one can attribute to each gene copy a value, the fitness asif $1 + p_i(-C + RB)$, such that the change of allele frequency in the population due to selection proceeds *as if* individuals were changing their behaviour to increase their fitness asif (called "inclusive fitness" by Hamilton 1964). Here, $-C + RB$ can be recognized as being the *average effect* of a mutant allele (Falconer and Mackay 1996; Lynch and Walsh 1998; Frank 1997), so that Hamilton's construction actually holds even in the presence of non-additive gene action. But it provides exactly the same information about the operation of natural selection and the behavior of individuals as the Price equation itself: selection favors those alleles that are associated to the fitness of their carriers. Since there is no general, nor even univocal, relationship between the change in fitness and

allele frequency change under natural selection, any individual-as-maximizing-agent analogy is constrained to depend on specific assumptions.

As illustrated above, such an analogy can be identified in the absence of social interactions (Grafen 2002), and for social interactions between relatives when phenotypic effects are additive separable. But owing to a lack of dynamic sufficiency we failed to find a satisfying proof of this later case in Grafen's (2006) writings. When phenotypic effects are not additive separable (the rule when social interactions occur), an individual maximand that is formally maximized in an ideal evolutionary process (trait substitution model and additive gene action) can be constructed. But no clear individual-as-maximizing-agent analogy emerges unless further assumptions are made, like that of Pareto optimal evolutionary stable states (characterized by $\max_{z_i,z_j}\left[f(z_i,z_j)+f(z_j,z_i)\right]$ when individual i and j interact), and where individuals can then be regarded as fecundity (or group fecundity) maximizers. This piecemeal identification of maximands makes us skeptical of the importance in evolutionary biology of strict individual-as-maximizing-agent analogies (i.e., strict optimization as opposed to concepts of best-responses). But for all the maximands encountered in this paper, their derivative is proportional to Hamilton's inclusive fitness effect. This describes the direction of selection under all conditions, a general message worth recalling.

Acknowledgments This work was partly funded by Swiss NSF Grant PP00P3-123344. We thank Christine Clavien, Alan Grafen, Charles Mullon, and Samir Okasha for useful comments on various drafts.

References

Bourke A (2011) Principles of social evolution. Oxford University Press, Oxford
Champagnat N, Ferrière R, Méléard S (2006) Unifying evolutionary dynamics: from individual stochastic processes to macroscopic models. Theor Popul Biol 69:297–321
Charlesworth B (1978) Some models of the evolution of altruistic behaviour between siblings. J Theor Biol 72:297–319
Charnov EL (1976) Optimal foraging, the marginal value theorem. Theor Popul Biol 9:129–136
Crow JF, Kimura M (1970) An introduction to population genetics theory. Harper and Row, New York
Dawkins R (1982) The extended phenotype. Oxford University Press, Oxford
Dieckmann U, Law R (1996) The dynamical theory of coevolution: a derivation from stochastic ecological processes. J Math Biol 34:579–612
Eshel I (1983) Evolutionary and continuous stability. J Theor Biol 103:99–111
Ewens WJ (2004) Mathematical population genetics. Springer, New York
Ewens WJ (2011) What is the gene trying to do? Br J Philos Sci 62:155–176
Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics, 6th edn. Longman, Essex
Fisher RA (1930) The genetical theory of natural selection. Clarendon Press, Oxford
Frank SA (1997) The Price equation, Fisher's fundamental theorem, kin selection, and causal analysis. Evolution 51:1712–1729
Frank SA (1998) Foundations of social evolution. Princeton University Press, Princeton
Gillespie JH (2004) Population genetics: a concise guide. Johns Hopkins, Baltimore
Grafen A (1999) Formal Darwinism, the individual-as-maximizing-agent analogy and bet-hedging. Proc R Soc Lond Ser B Biol Sci 266:799–803
Grafen A (2000) Developments of the Price equation and natural selection under uncertainty. Proc R Soc B Biol Sci 267:1223–7
Grafen A (2002) A first formal link between the Price equation and an optimization program. J Theor Biol 217:75–91

Grafen A (2006) Optimization of inclusive fitness. J Theor Biol 238:541–563
Grafen A (2007a) The formal Darwinism project: a mid-term report. J Evol Biol 20:1243–1254
Grafen A (2007b) An inclusive fitness analysis of altruism on a cyclical network. J Evol Biol 20:2278–2283
Grafen A (2008) The simplest formal argument for fitness optimization. J Genet 87:421–33
Grafen A (2014) The formal darwinism project in outline. Biol Philos 29(2). doi:10.1007/s10539-013-9414-y
Hamilton WD (1964) The genetical evolution of social behaviour, 1. J Theor Biol 7:1–16
Hamilton WD (1970) Selfish and spiteful behavior in an evolutionary model. Nature 228:1218–1220
Johnstone RA, Woodroffe R, Cant M, Wright J (1999) Reproductive skew in multimember groups. Am Nat 153:315–331
Kingman J (1961) A mathematical problem in population genetics. Math Proc Camb Philos Soc 57:574–582
Lehmann L (2007) The evolution of trans-generational altruism: kin selection meets niche construction. J Evol Biol 20:181–189
Lehmann L (2012) The stationary distribution of a continuously varying strategy in a class-structured population under mutation-selection-drift balance. J Evol Biol 25:770–787
Lessard S, Ladret V (2007) The probability of fixation of a single mutant in an exchangeable selection model. J Math Biol 54:721–744
Lion S, Gandon S (2009) Habitat saturation and the spatial evolutionary ecology of altruism. J Evol Biol 22:1487–1502
Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sinauer, Massachusetts
Mas-Colell A, Whinston MD, Green JR (1995) Microeconomic theory. Oxford Unversity Press, Oxford
Matsuda H, Abrams PA (1994) Runaway evolution to self-extinction under asymmetrical competition. Evolution 48:1764–1772
Maynard Smith J (1982) Evolution and the Theory of Games. Cambridge University Press, Cambridge
Maynard Smith J, Szathmary E (1995) The major transitions in evolution. Oxford University Press, Oxford
McNamara J, Houston AI, Collins EJ (2001) Optimality models in behavioral ecology. SIAM Rev 43:413–466
Metz JAJ, Geritz SAH, Meszéna G, Jacobs FJA, van Heerwaarden JS (1996) Adaptive dynamics: a geometrical study of the consequences of nearly faithful reproduction. In: Strien SJ, VerduynLunel SM (eds) Stochastic and spatial structures of dynamical systems. North-Holland, Amsterdam, pp 183–231
Metz JAJ, Mylius SD, Diekmann O (2008) When does evolution optimize. Evol Ecol Res 10:629–654
Moran PAP (1964) On the nonexistence of adaptive topographies. Annals of Human Genetics 27:383–393
Mylius SD, Diekmann O (1995) On evolutionarily stable life histories, optimization and the need to be specific about density dependence. Oikos 74:218–224
Ohtsuki H, Hauert C, Lieberman E, Nowak MA (2006) A simple rule for the evolution of cooperation on graphs and social networks. Nature 441:502–505
Parker GA, Maynard Smith (1990) Optimality theory in evolutionary biology. Science 349:27–33
Price GR (1970) Selection and covariance. Nature 227:520–521
Price GR (1972) Fisher's "fundamental theorem" made clear. Ann Hum Genet 36:129–40
Reeve HK (1989) The evolution of conspecific acceptance thresholds. Am Nat 133:407–435
Rousset F (2003) A minimal derivation of convergence stability measures. J Theor Biol 221:665–668
Rousset F (2004) Genetic structure and selection in subdivided populations. Princeton University Press, Princeton
Stearns S (1992) The evolution of life histories. Oxford University Press, Oxford
Taylor PD (1989) Evolutionary stability in one-parameter models under weak selection. Theor Popul Biol 36:125–143
Wenseleers T, Gardner A, Foster KR (2010) Social evolution theory: a review of methods and approaches. In: Szekely T, Moore A, Komdeur J (eds) Social behaviour: genes, ecology and evolution. Cambridge University Press, Cambridge, pp 132–158
Wright S (1942) Statistical genetics and evolution. Bull Am Math Soc 48:223–246